# Sub-Annual to Annual Dynamics of Alaskan Ice-Marginal Lakes from Automated Image Classification Using Google Earth Engine

by
Anthony M. Hengst

Honors Thesis

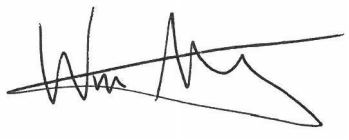
Appalachian State University
Submitted to the Department of Geological and Environmental Sciences
and The Honors College

in partial fulfillment of the requirements for the degree of

Bachelor of Science

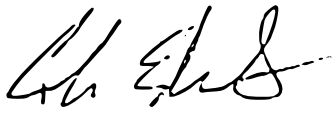
December, 2020

Approved by:



William Armstrong, Ph. D.

Song Shu, Ph. D.



Cole Edwards, Ph. D.

Jefford Vahlbusch, Ph. D., Dean, Honors College

# Abstract

Ice-marginal lakes play an important role in glacier dynamics and downstream hydrology. Proglacial lakes may alter glacial mass loss by enabling submarine melt and by providing a body of water into which glaciers may calve, and provide a basin which traps glacial sediment. Ice-dammed lakes play a critical role in the generation of outburst floods and must be monitored for human safety in downstream environments.

Observation of ice-marginal lakes from satellite imagery provides valuable insight into these remote systems because in-situ data are difficult to obtain over a large study area. However, even large-scale remote sensing of these lakes is difficult due to their varied spectral appearance and the complex interface between sediment-laden, iceberg filled lakes and their adjacent crevassed and water-covered glaciers.

Previous remote sensing studies feature coarse temporal sampling of lake behavior over a multi-decadal timescale. We seek to investigate how ice-marginal lakes evolve over sub-annual to annual timescales. Ice-marginal lakes are intimately connected to glacial systems, which can vary over seasonal cycles and longer-term cycles in the case of some surging glaciers. We develop a robust remote sensing method to provide observations of ice-marginal lakes across Alaska, a region whose ice-marginal lakes have received comparatively little attention.

We develop an automated routine implemented in Google Earth Engine to investigate short-term glacial lake area changes across southern Alaska over the Landsat 8 era (2013-present). We create monthly estimates of ice-marginal lake area by applying a supervised Mahalanobis minimum-distance land cover classifier to Landsat 8 imagery. We optimize image processing parameters by running a suite of classifications and selecting the parameters that minimize error against a set of manually-delineated lakes and achieve an F-score from 0.33 in the most challenging test regions to 0.77 at best.

In an exploration using Monte Carlo simulations, we interrogate our data to characterize the uncertainty in lake area associated with sparse temporal sampling. These data provide short-term context for multi-decadal studies and yield insight into the uncertainty inherent in remote-sensing studies of ice-marginal lake.

# Introduction

## Ice-marginal Lakes

Ice-marginal lakes are lakes which share a margin with glacial ice (*Tweed and Carrivick*, 2015). Broadly, there are two types of ice-marginal lakes: proglacial and ice-dammed (Figure 1). Ice-marginal lakes are dynamic features which exhibit both short- and long-term patterns of behavior (*Post and Mayo*, 1971; *Carrivick and Tweed*, 2013).

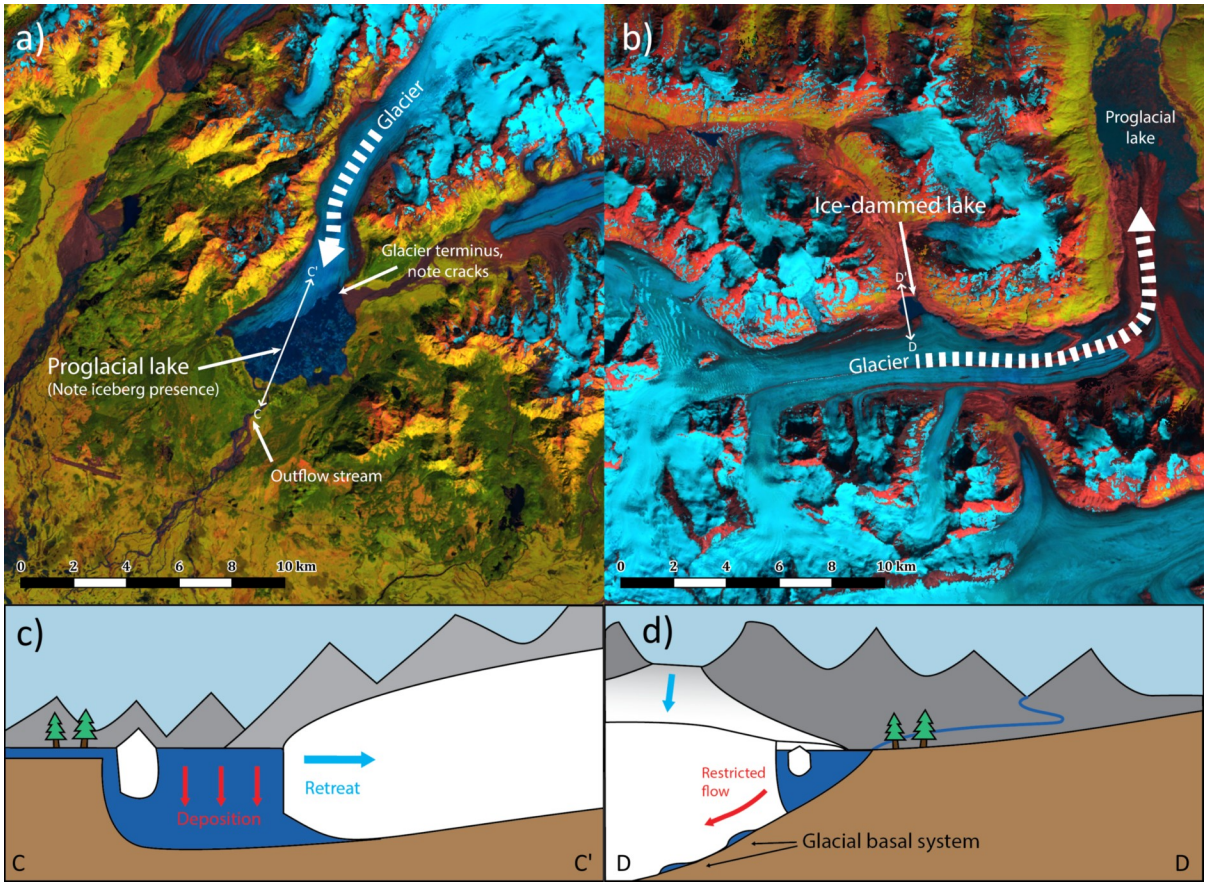



Figure 1: Ice-marginal lakes from space with idealized cross-sections. A proglacial lake in Landsat band 6-5-4 falsecolor imagery (a). An ice-dammed lake in the same falsecolor imagery (b). (c) Idealized cross-section of a proglacial lake, viewed from the side of the glacier. The space it occupies is the overdeepening left behind by glacier retreats. (b) Idealized cross-section of an ice-dammed lake, viewed looking up the glacier slope. It cannot drain through the glacier basal system due to the weight of the ice dam impounding it.

Proglacial lakes are are found at the glacier terminus occupying the valley left by glacier retreat (*Carrivick and Tweed*, 2013; *Bogen et al.*, 2015; e.g. Figure 1a,c). Proglacial lakes provide a body of water into which a glacier may calve (*Walder et al.*, 2006; *Larsen et al.*, 2015; *Chernos et al.*, 2016; *Otto*, 2019). This body of water also allows the glacier to lose mass by submarine melting (*Tsutaki et al.*, 2011; *Otto*, 2019). These effects have a negative impact on glacial mass balance (*Larsen et al.*, 2015), but the magnitude of the impact is unclear (*Larsen et al.*, 2015). Further, proglacial lakes trap sediment that would otherwise be transported through outflow streams into the down-glacier environment (*Bogen et al.*, 2015; *Otto*, 2019), forming an important regulator for down-stream ecosystems (*Dorava and Milner*, 2000).

Ice-dammed lakes exist up-glacier, and are impounded by glacier ice (*Tweed and Russell*, 1999; *Carrivick and Tweed*, 2013; *Otto*, 2019; e.g. Figure 1b,d). These lakes are frequently the

source of glacial outburst floods when they overcome the ice dam and outflow through the glacial basal system (*Post and Mayo*, 1971; *Tweed and Russell*, 1999; *Carrivick and Tweed*, 2013; *Jacquet et al.*, 2017; *Falatkova et al.*, 2019; *Otto*, 2019), so understanding their dynamics is important for flood hazards preparedness.

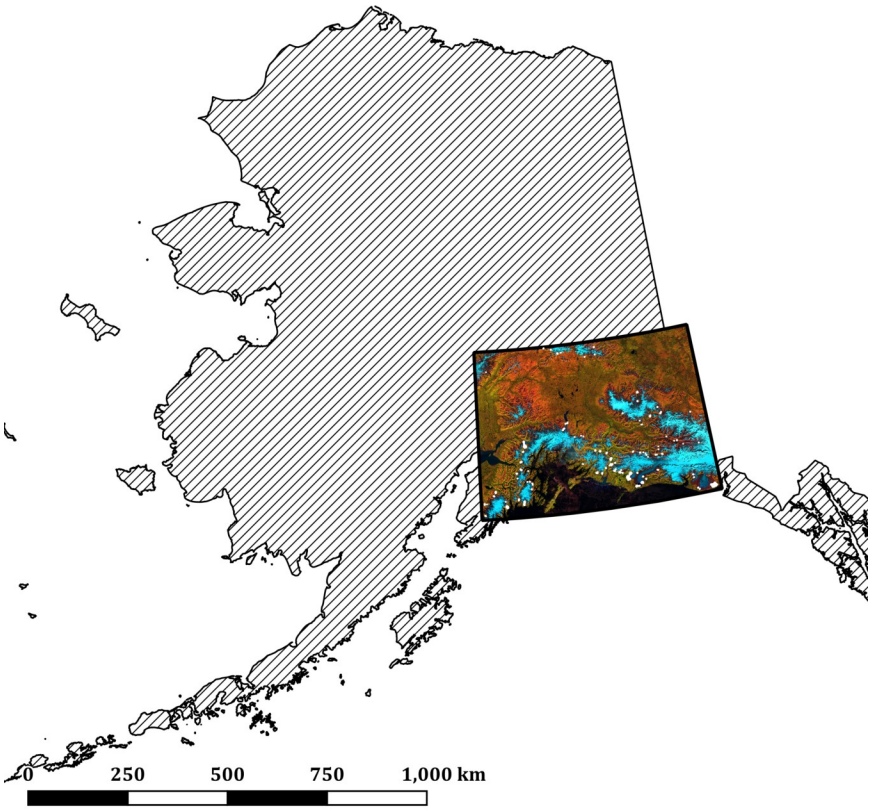



Figure 2: Study area, including much of the Alaska and Chugach ranges.

We conduct our study in south-eastern Alaska (Figure 2). The ice-marginal lakes of this region have received less attention than those of the Himalayas, but it has be found to generally be losing glacial mass (*Larsen et al.*, 2015). While globally, climatic deglaciation (*Kargel et al.*, 2014) has driven increases in ice-marginal lake number and volume (*Tweed and Carrivick*, 2015; *Shugar et al.,* 2020), the connection between deglaciation and ice-marginal lakes is less well-defined in Alaska, with some studies suggesting that the number of ice-dammed lakes have become less common in recent years (*Wolfe et al.*, 2014). Our study of this region seeks to both map Alaskan ice-marginal lakes and characterize the impact of their short-term variability on studies of their long-term behavior.

Both types of ice-marginal lakes are closely connected to the basal hydrology of their adjacent glacier (*Tweed and Russell*, 1999; *Falatkova et al.*, 2019), and are fed primarily by glacial melt in the forms of surface runoff and basal flow (*Carrivick and Tweed*, 2013). This connection to the variable glacial hydrologic system enables their rapid change in volume and area, a characteristic trait (*Carrivick and Tweed*, 2013; *Otto*, 2019). Due to their rapid temporal variability and the difficulty of obtaining in-situ measurements in alpine terrain (*Larsen et al.*, 2015; *Jacquet et al.*, 2017), remote sensing of ice-marginal lakes is an essential tool for studying these lakes (*Shugar et al.*, 2020).

While lake area can be obtained from remote sensing data, there is no way to determine lake volume where lake depth data are absent (*Huggel et al.,* 2002). However, lake area is strongly correlated to lake volume (*Huggel et al.,* 2002; *Cook and Quincey*, 2015), so we use it as a proxy for volume in this study.

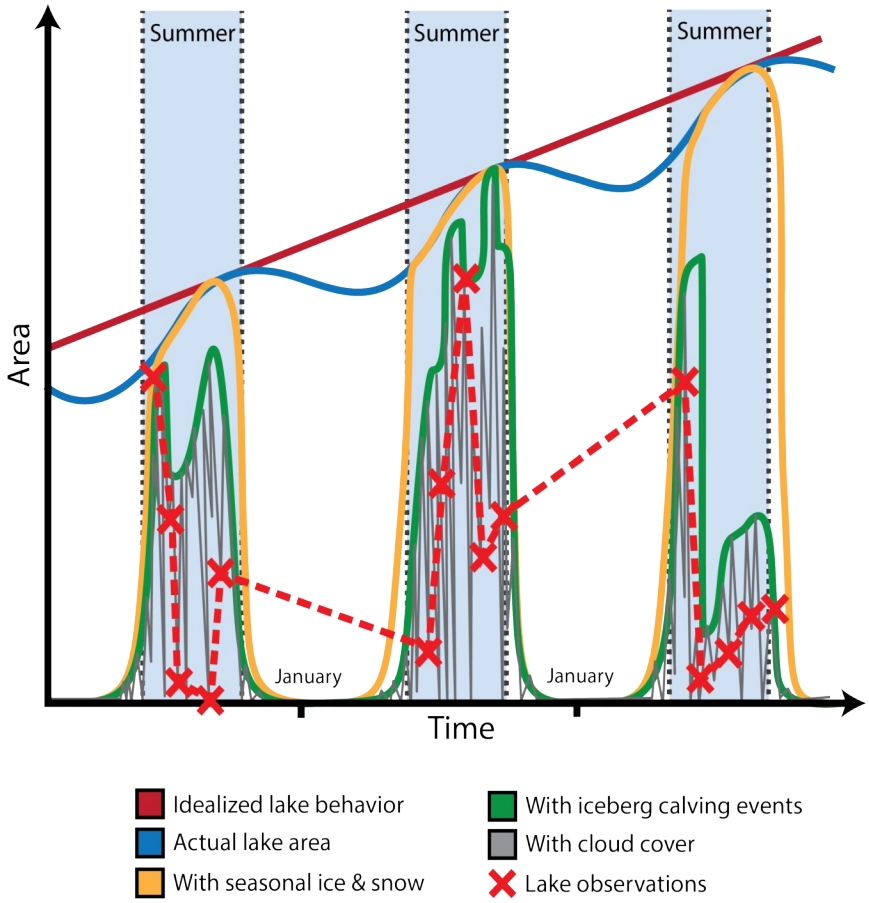



Figure 3: Theoretical limitations of remote sensing of glacial lakes. The actual lake area (blue) is obscured seasonally by winter ice & snow (yellow). Additionally, iceberg calving obscures the water surface (green), and cloud cover further obscures observations on a daily or hourly basis (gray). Even the most accurate classifier (red marks) may perform poorly due to ice and cloud cover.

There are conceptual limitations on how well a remote sensing study can observe glacial lakes. Consider such a lake growing in area over several years (Figure 3, blue line). While we wish we could continuously observe it, it will be obscured every winter by ice and snow (yellow). Further, even in the summer calving events will prevent some area from detection by iceberg cover (green). Finally, clouds will frequently obscure the lake partially or entirely (gray). These issues make accurate delineation of ice-marginal lake change from remotely-sensed data a challenge, and motivate our use of high temporal resolution data. This is a data-heavy task, and we turn to Google Earth Engine to help mitigate the impact of the data load.

## Google Earth Engine

Google Earth Engine is a cloud-computing service provided by Google which allow us to process very large amounts of remote sensing data without having to download input or intermediate data. Additionally, tasks may be submitted to the cloud servers to run in parallel. We use tasks which deposit output rasters in an associated Google Drive account. Our methodology is both data- and processing-heavy, and use of Google Earth Engine greatly streamlined our workflow.

Figure 4 will illustrate which portions of our workflow were completed in a single Google Earth Engine script.

## Overview of Distance Classifiers

Observation of lakes from remote sensing data is a *classification problem*, as the land cover type contained by each pixel must be classified as water or not-water. Distance classifiers have been applied to remote sensing land cover classification problems since at least the 70s (*Wacker and Landgrebe*, 1972). We provide a brief overview of the simplest of these trusted methods and then discuss the particular type of distance classifier we have chosen, a Mahalanobis distance classifier (*Gong and Howarth,* 1990).

To perform classification of raster data, a distance classifier dimensionalizes each pixel in spectral space and calculates the distance between each pixel and the centroid of possible classes to which it may belong (*Parr and Schucany*, 1980). In our application, we have three spectral bands under consideration (SWIR, NDVI, and band 8 local variance (B8LVR)), so each pixel may be envisioned as positioned in a 3-dimensional space with axes described by those bands. A distance classifier describes regions in spectral space which characterize pixel classes (for example, water is dark in all of SWIR, NDVI, and B8LVR, so the water classes are defined as regions near to (0,0,0) in spectral space), and pixels are classified according to distance to these class regions. A minimum-distance assigns classes to all input pixels by selecting the class to which the pixel is closest—the minimum distance.

A classifier may be untrained, in which case it determines the position of class centroids in spectral space according to some optimization function. A trained classifier is assigned classes and training data (typically, a collection of pixels) which have been manually chosen to represent a class. From this training data, the classifier will construct the training region by statistically characterizing the training data's distribution and position in spectral space.

Minimum-distance classifiers differ based on how they measure distance. Euclidian minimum-distance classifiers are the easiest to understand, as they utilize the most commonly-understood measure of distance—the square root of the square of the differences in each spectral dimension. This familiar metric is how distances are usually measured between points in flat space. In a Euclidian distance classifier, class regions are single points in spectral space—typically, the mean of the training data for that class.

The Mahalnobis measure of distance is not a vector norm (distance between two points), as is the Euclidian distance (*McLachlan*, 1999). Rather, it is a measure of distance from a distribution to a point. It measures how many standard deviations away from the mean of a distribution a point lies, with consideration to the orientation of the principle axes of the distribution and the possibility that deviation may vary along them. Using the Mahalnobis distance in a minimum-distance classifier partially mitigates some issues associated with asymmetrical class distributions as well as differences between highly specific classes (e.g.: ocean water), which have a very small footprint in spectral space, and broad classes (e.g.: vegetation): which can encompass a large area in spectral space.

While more sophisticated classifiers exist, the Mahalnobis minimum-distance classifier is the best solution available to us considering limitations of Google Earth Engine, and results of Mahalanobis minimum-distance classifiers strongly resemble those of more sophisticated maximum-likelihood classifiers (*Gong and Howarth*, 1990).

# Methodology

Below, we describe our method using Landsat false-color images from 2013 through 2019 to automatically identify and delineate lakes boundaries. We clean these lake boundaries manually, and use the clean, accurate data in Monte Carlo simulations to provide estimates of lake area change rate uncertainty for poorly-observed ice-marginal lakes.

This workflow is long and complex, so we provide a summary flowchart in Figure 4.

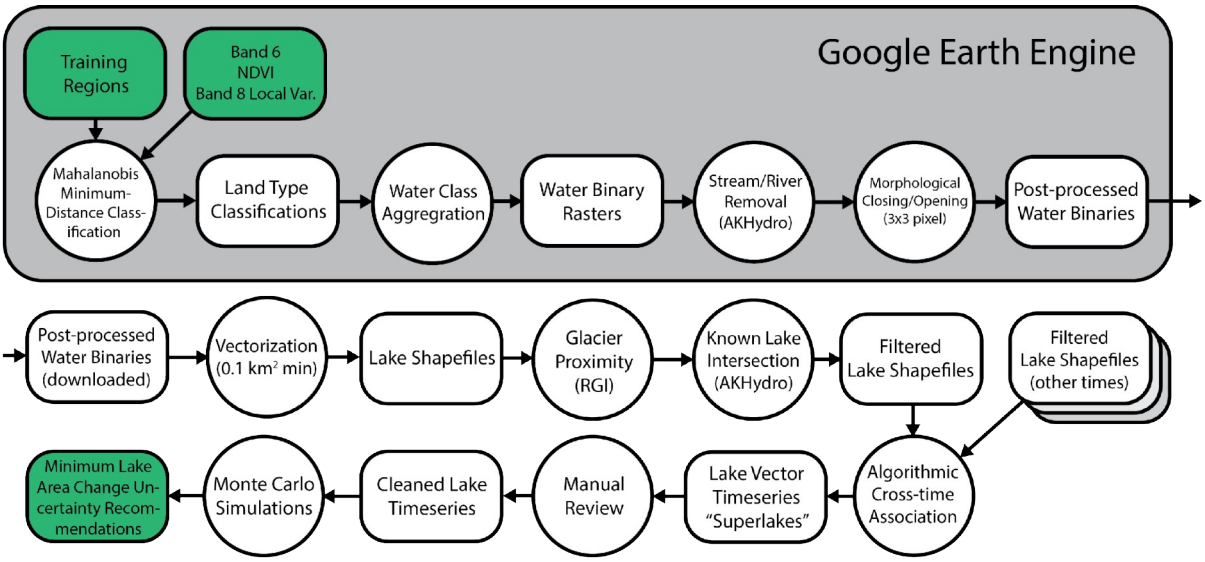



Figure 4: Methodology workflow.

## Selection of Training Data

From a Landsat 8 three-month composite (July through September, 2018), we hand-delineate regions in our study area (Figure 2) representative of the following fourteen classes: accumulation zone ice and snow, accumulation zone ice and snow in shadow, bare glacial ice, wet supraglacial debris, dry supraglacial debris, dark wet soil, bare rock, yellow (in Landsat band 6-5-4 falsecolor) land, green land, dark green land, blue lake water, non-glacial lake water, and proglacial lake water. Fifty pixels (a relatively low number due to limitations in the Google Earth Engine classifier implementation) are randomly sampled from each region (of which there may be multiple per class) and provided to the classifier as training data.

We define multiple similar classes, such as the different vegetation-related land cover types, because while minimum-distance classifiers may be able to distinguish class distribution anisometries, it still assumes members of a class are distributed normally. Thus, non-"ellipsoidal" classes, or other classes where the majority of elements are not near the mean, will be improperly classified. Additionally, when such classes are very large, although they may be well-defined (consider a crescent shape), standard deviations calculated relative to the mean may be very large. Thus, Malahnobis distances calculated from such distributions will not be representative of membership within the distribution. Breaking potentially expansive, non-normal classes into multiple smaller classes improves the quality of the classification using minimum-distance classifiers.

## Producing Input Raster Data

In each year, we produce on Landsat 8 composite image for each month from June through September over our study area (Figure 2) using the cloud-based Google Earth Engine (GEE). We

utilize the GEE simpleComposite() function to mosaic images in such a way as to minimize cloud cover in the final composite.

Based on this composited Landsat 8 imagery, we form false-color images by setting band 6 (the reflectance of short wave infrared band (SWIR)) to red, the Normalized Difference Vegetation Index (NDVI) to green, and the local variance of band 8 (the reflectance of panchromatic band) to green.

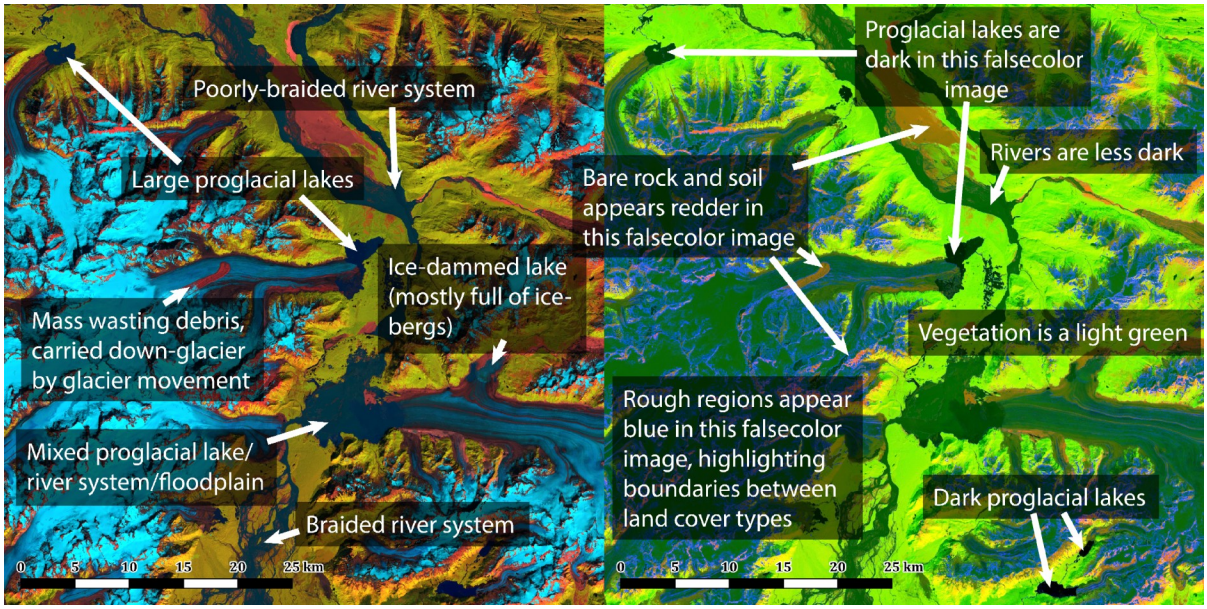



Figure 5: False color rasters. On the left is a standard band 6-5-4 false color visualization. On the right is shown the inputs into our classifier, with B6, NDVI, and B8LVR mapped to red, green, and blue, respectively.

The SWIR reflectance is of interest to us because soil and exposed rock is reflective in this band while water is very absorptive, making SWIR a useful band for differentiating water from non-water land cover types.

NDVI describes the relationship between red visible reflectance and near-infrared reflectance of land cover (*Pettorelli*, 2013). It is tuned to return a positive value in the presence of vegetation, a zero value for bare ground, and a negative value for water. Healthy vegetation may be characterized by low reflectance in the visible red wavelengths due to the absorptive properties of chloroplasts and high reflectance in the infrared wavelengths due to the spectral characteristics of plant cell walls (*Pettorelli*, 2013). The NDVI is commonly used to characterize the quantity and health of crop or forest cover from satellite data. While water, not vegetation, is our land cover type of interest, we find the NDVI is a better discriminator than the normalized difference water index (NDWI) in distinguishing ice-marginal lakes from spectrally similar land cover types.

Local variance of panchromatic optical reflectance (band 8 local variance, B8LVR) allows us to differentiate between spectrally similar but texturally different (smooth vs. rough) land cover types. We compute the standard deviation of reflectance in a 3 by 3 pixel window as a measure of a surface's optical roughness. On Landsat 8 imagery, this is a ninety by ninety meter region. A surface with a homogeneous appearance will have low local variance, while a heterogeneous surface (e.g., crevassed glacier, variably rocky hillslope) will have high local variance.

## Raster Post-processing and Vectorization

We obtain a binary raster map of water identifications by aggregating all water classes into a single "water" supergroup. This is exported from Google Earth Engine through Google Drive. We reduce pixel-level noise using morphological operations. We remove noise by a 3×3 morphological opening process (erosion followed by dilation). Next, we close small holes in our water identifications by a 3×3 morphological closing process (dilation followed by erosion).

We exclude all pixels which are above a 10° slope threshold, using elevation data from the ALOS Global Digital Surface Model (*Tadono et al.*, 2014), produced and distributed by JAXA. Removing sloped regions of the study area from consideration allows us to focus on our land cover type of interest—the flat surfaces of lakes.

To exclude water classifications which are valid but not of interest to us, we use a data product known as AKHydro which has a hand-delineated "streams" product to remove all pixels which may be in a river or floodplain from the binary map (Figure 6). AKHydro is a National Hydrography Dataset produced and distributed by the Alaska Center for Conservation Science (http://akhydro.uaa.alaska.edu/data/nhd/). These water objects (lakes, streams, etc.) are highly accurate but do not represent any particular time, as the data product is irregularly and incompletely updated.

After post-processing, we vectorize this binary raster to generate shapefiles for water regions. After we vectorize the binary map, we obtain vector objects representing our non-stream water bodies. Hereinafter, we will call such vector objects lakes.

## Vector Object Post-processing

During the vectorization process, we reject lakes smaller than 0.1 $km^2$ in area. Landsat 8 SWIR imagery consists of 30 meter by 30 meter pixels. At this scale, 0.1 $km^2$ is a region slightly larger than 11 pixels. It is very difficult to positively distinguish features smaller than this even with human oversight. Many "water features" smaller than the threshold are results of noise, bad imagery, or poor classification, and removing them early in our processing method dramatically reduces the volume of objects under consideration, improving script runtimes correspondingly.
Some real water features are below the size threshold will be excluded from this study, but their signal-to-noise ratio is too low to be useful for observing processes of lake area change. At a size of only about 11 pixels, the smallest observable change (1 pixel) will produce 9% variation in lake area. Any misclassification—even of a single pixel—will further obscure real changes of lakes of this scale.

To reduce false lake positives and focus our study on specifically ice-marginal lakes, we filter the lake shapefiles by lake area, proximity to Randolph Glacier Inventory (RGI) glaciers, and intersections with AKHydro lakes. The RGI is a spatial data product consisting of a global inventory of glacier boundaries, produced and distributed by the Global Land Ice Measurements from Space (GLIMS) team (*Pfeffer et al.*, 2014). The AKHydro data product is discussed in the section above.

We use the AKHydro lake filter as a preliminary filtering method which ensures we are looking at lake classifications only where water exists without compromising our capacity to accurately delineate temporally varying lake margins. We discard all identifications of water which do not intersect an AKHydro lake and retain all which do. This filtering process reduces the chance that incorrect positive classifications (wet ice, dark shadows) will be retained.
Lakes of interest are ice-marginal. We remove non-ice-marginal lakes by only accepting lakes within 3 km of a glacier in the Randolph Glacier Inventory. Further manual processing is required to remove lakes that are close but not in direct contact with glaciers.
The lake boundaries obtained from later images are algorithmically associated with those from prior images to form a three-dimensional (two spatial, one temporal) lake object, a collection of

lake boundaries which represents the evolution of a physical lake through time. This association is made by identifying those lake boundaries which overlap, even if they are from different points in time. This process allows a lake which is fragmented at a given point in time (identified in multiple vectors), to be reconciled and grouped into a single vector object per point in time regardless of fragmentation.

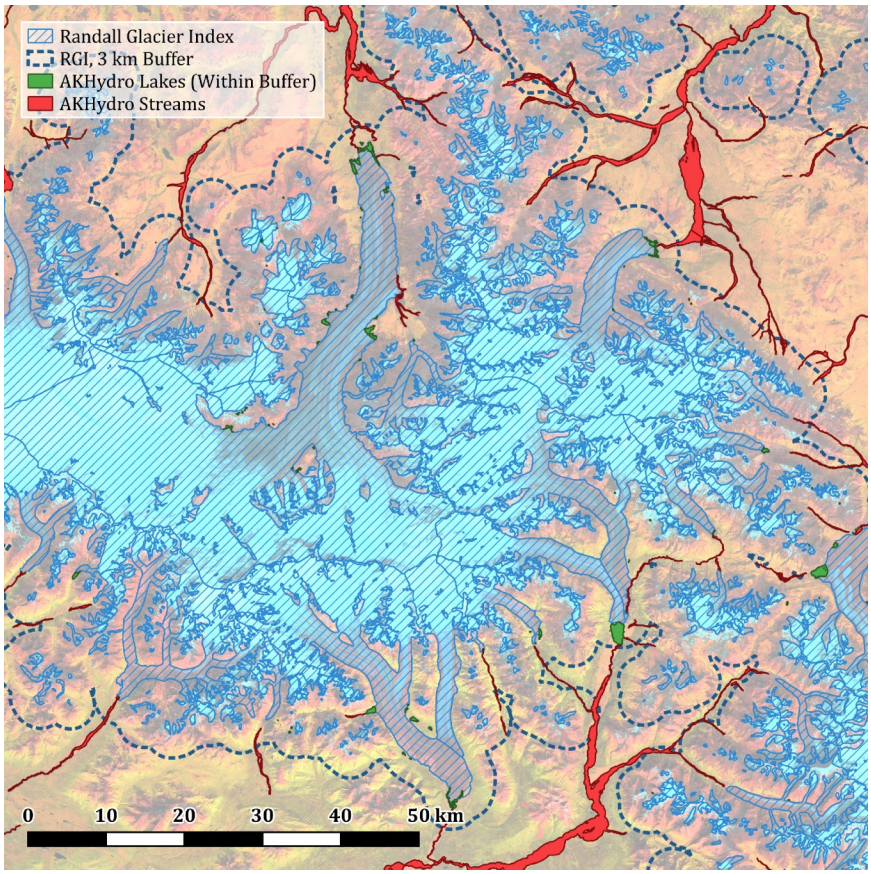



Figure 6: Post-processing spatial filters: buffered RGI, AKHydro lakes, and AKHydro streams.

## Methodology Decision Process

We seek to optimize our classification method and raster post-processing routine for accurate identifications of ice-marginal lakes, potentially at the expense of accurate identifications of other land cover types. While we have fourteen different land cover classes ranging from glacial ice to floodplain, we are only interested in the three water classes. To optimize our classification method, we embarked on an exploratory process in which we adjusted input parameters to the classifier and evaluated the accuracy of each result against previous manually delineated ground-truth lakes in the test regions of our study area. Those input parameters which produced the best results we selected for our study.

To ensure representation of all types of ice-marginal features, we selected three different test regions (Figure 7a. Region 1 contained large, ice-free proglacial lakes with streams connecting them (Figure 8a). Region 2 contained intermediate-size proglacial lakes and ice-dammed lakes (Figure 8b). Region 3 contained very iceberg-choked proglacial lakes and many small supraglacial lakes (Figure 8c). While ultimately, evaluation of classifier quality was done by manual assessment of the produced shapefiles, we used a metric known as the F-score to characterize the classifier accuracy and accelerate evaluation of the quality of a given set of

input parameters. The F-score, or F-measure, is a performance metric commonly used for qualifying the power of binary classification methods (*Dembczynski et al.*, 2011), which ranges from a value of 0 at worst and 1 at best.

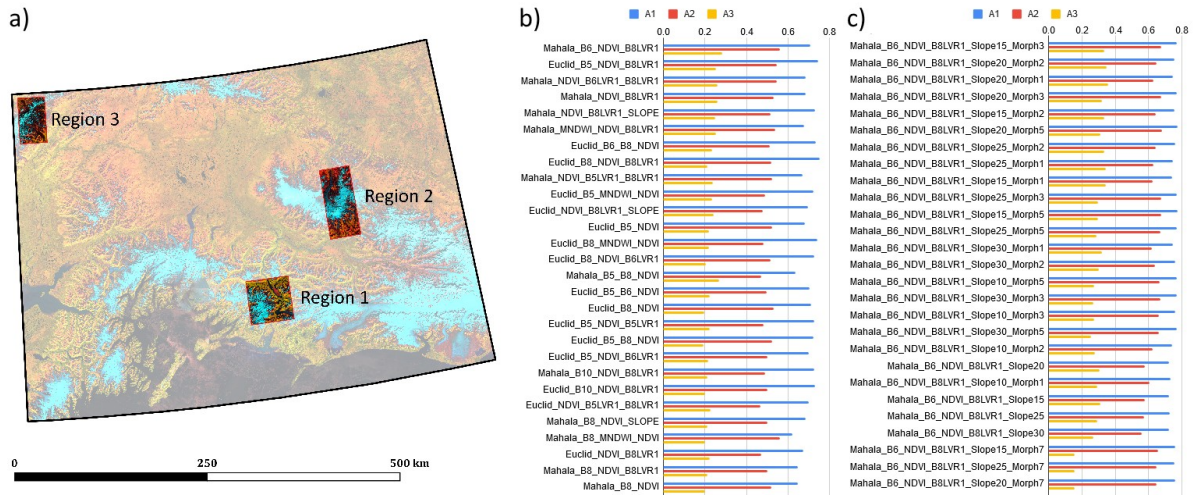



Figure 7: (a) Location of three test regions within our study area. (b-c) Evaluation and selection of classifier details by F-score comparison. F-score performance in areas 1, 2, and 3 are indicated by legend colors. Bar charts are incomplete for size. Initially (b), we permuted input data and classifier distance measure, and then (c) permuted maximum slope threshold and morphological kernel size of the best-performing classifier from part (b). The overall best-performing classifier was selected.

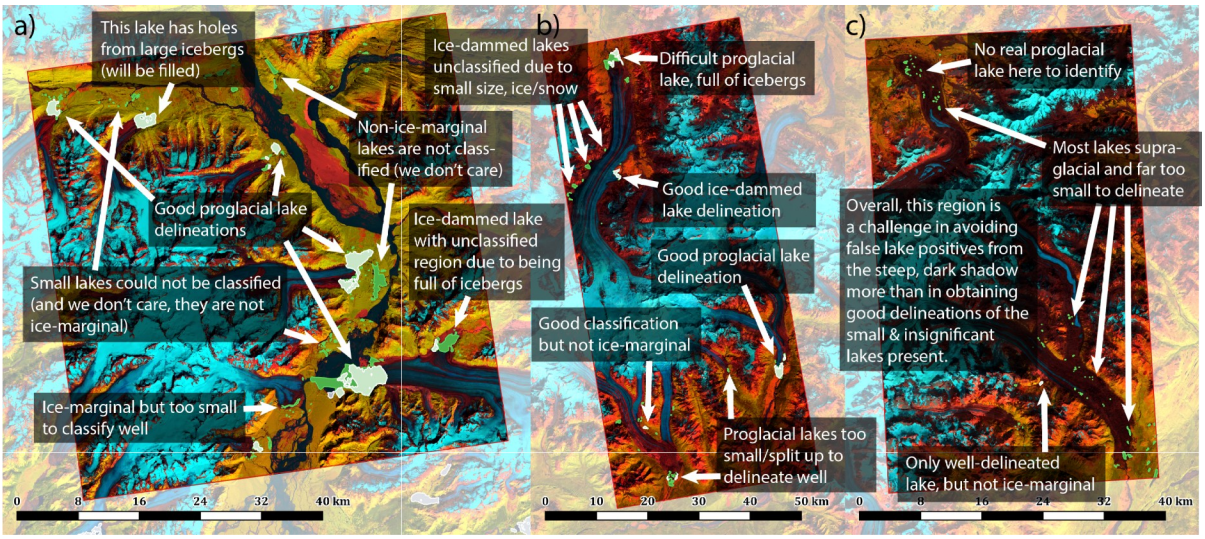



*Figure 8: Detail of each test region and comparison of final classification results (white) with ground-truth (green).*

We evaluated hundreds of different classifiers (Figure 7b-c), varying selection of input bands (Landsat 8 bands 5, 6, 8, 10), spectral difference indices (NDVI, modified normalized difference water index (MNDWI), *Xu* (2006)), local variance maps (for bands 5, 6, and 8), different slope thresholds (1°-30° cutoff), and difference kernel sizes for morphological post-processing operations (1-7 pixel square kernel size). To reduce our search space and prevent from building an incomprehensible black box, we limited our search to combinations of three or fewer data sources, as this reduced the possible combinations by orders of magnitude and allowed us to visualize the input data as an RBG false color (e.g.: Figure 5a).

Ultimately, the combination of band 6, NDVI, and band 8 local variance inputs with a 15° slope cutoff and a 3×3 morphological kernel was decided upon as the best-performing for our application of water surface identification. This classifier performed with an F-score of 0.77 in the best-performing test region and 0.33 in the most challenging (Figure 7c, top row). To ensure our statistical measure of performance represented fidelity to ground-truth, we manually reviewed the classification results of the best few classifiers in comparison to the manually-delineated lakes (Figure 8).

## Manual Data Review

The process described above identified 216 lakes in our study area with 2676 monthly observations total, an average of 12.3 observations per lake. Our classified data, while accurate, is imperfect, and there is error in some lake boundaries. This error is from several sources. We manually review each monthly observation of these lakes and manually flag it as good or bad, according to the source of the error.

While manual review is both time-consuming and subjective, it is the best way to remove erroneous observations after automated classification. Because we ultimately seek to study the uncertainty in lake area change calculations caused by low observation rates, we must minimize classification, imagery, and obstruction error in our data lest it obscure the effect we seek to study. We clean the data by manually winnowing good observations from bad ones. If error is demonstrated in the lake area delineated, we will make a decision on each lake regarding whether it and its good delineations ought to be included in our dataset.

## Monte Carlo Modeling

Ice-marginal lakes can exhibit major short term variation in area (*Post and Mayo*, 1971; *Carrivick and Tweed*, 2013; *Otto*, 2019). In studies which sample on annual or longer timescales, this short-term variation may obscure the rate of long-term area change in the lake. However, long-term remote sensing studies of ice-marginal lake area change may not have considered this factor. We seek to characterize the *lake area change uncertainty* associated with low sampling rates.

We do so by randomly resampling a subsample from our good data in a simulation of a more limited study. For example, consider an extreme example—a lake which mostly drains twice each summer (Figure 9a). Such sub-annual variation is physically improbable in an ice-marginal lake, but considering extremes is useful when trying to conceptually understand how lake area change uncertainty arises from low sampling rates. Now consider a limited study which is only able to observe this lake twice, once each in consecutive years.

Such a study may observe the lake in a number of ways (red, green, yellow dashed lines and markers, Figure 9a), without being any wiser to which derived estimate of lake area change is physically reasonable. Such a study could conclude the lake is dramatically losing area (red, Figure 9a-b), dramatically gaining area (green, Figure 9a-b), or is constant (yellow, Figure 9a-b). Because we have data on a sub-annual timescale, we are more aware of such short-term variations than more limited studies. If we simulate these studies, we can generate a distribution of these possible area change rates (Figure 9b). From this distribution we can calculate confidence intervals which characterize the uncertainty the lake would have in such a study. We can use our data as a base from which we can simulate such limited studies.

We run 100,000 Monte Carlo simulations per model on each lake to characterize the lake area change uncertainty associated with a lower sampling rate that a more limited study may have. This process results in a non-normal distribution of possible lake area change rates, given from the range of lake areas such studies *could* have observed. We then report the width of the 95% confidence interval, in percent area change per year.

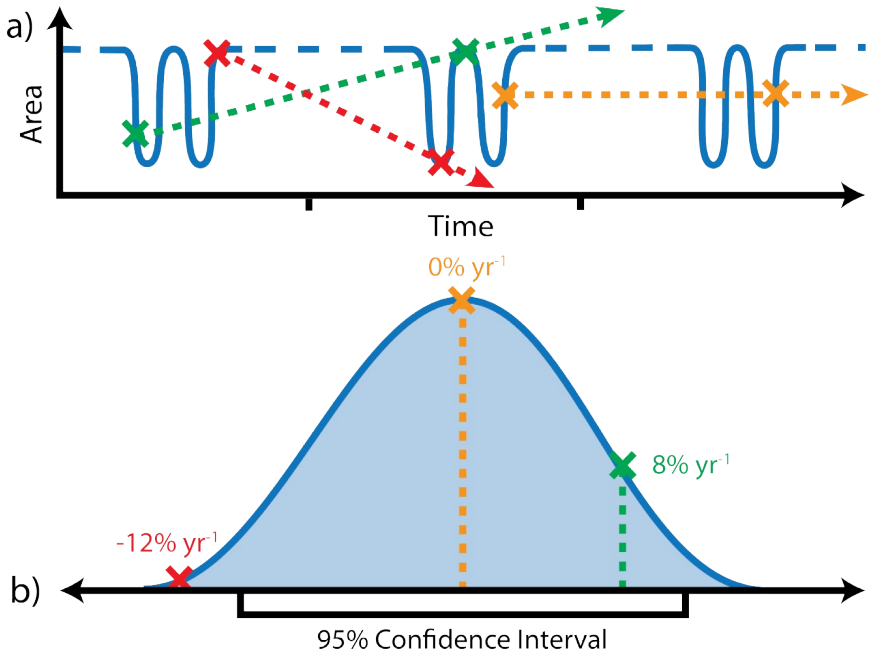



Figure 9: A conceptual model of a lake with sub-annual variation which may confound a study of its long-term behavior.

To mitigate the effect of slight errors in our own data and expand the number of possible resamplings, we resample each data point from a normal distribution where 95% of the possible resamplings lie within a ±5% range of our observed value. To prevent this from expanding the width of the 95% confidence interval we report, we run one set of Monte Carlo simulations as a control, using all our data but with the ±5% resampling around each data point. We subtract the width of this confidence interval from the width of the confidence interval of the set of Monte Carlo simulations which simulate limited studies, and report this value as the uncertainty associated with the lower sampling rate.

# Results

## Results from Manual Data Review

Table 1: Lake decision summary.

| | | Rejected | | | |
|---|---|---|---|---|---|
| | Total | Always Obscured | Poorly Classified | Lake Fragment | Accepted |
| *Proglacial* | 108 | 10 | 18 | 20 | 60 |
| *Ice-dammed* | 33 | 11 | 4 | 1 | 17 |
| *Not Ice-marginal* | 51 | 1 | 17 | 14 | 43 |
| ***Totals*** | 216 | 22 | 39 | 35 | 120 |

Of the lakes reviewed, 39 were rejected for not being well delineated enough at any given time (poorly classified), 22 were rejected for always being obscured due to cloud or ice cover (always obscured), and 35 were rejected for being a fragment of a lake which had not been attached to the rest of the lake (lake fragment). Sixty-six "lakes" were rejected due to being false positives (primarily due to misclassification of dark cloud shadow on a glacier where the AKHydro database had delineated a very small supraglacial or ice-dammed lake). One hundred

and twenty well-classified lakes were accepted for further analysis, and only accurate observations at each were retained.

Table 2: Results of the manual review. Observation codes, assigned by expert review.

| | | Discarded | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Obstruction Error | | Classification Error | | Imagery Error | |
| | **Total** | **Mostly Good** | **Ice Obs.** | **Cloud Obs.** | **False Neg.** | **False Pos.** | **Bad Imagery** | **Good** |
| *Proglacial* | 2160 | 183 | 466 | 573 | 338 | 76 | 44 | 480 |
| *Ice-dammed* | 773 | 30 | 348 | 196 | 62 | 24 | 8 | 105 |
| *Not Ice-marginal* | 1204 | 66 | 151 | 339 | 279 | 11 | 19 | 380 |
| ***Totals*** | 4137 | 279 | 965 | 1105 | 641 | 111 | 71 | 965 |

These decisions on lake quality were informed by manual review of 4137 individual lake observations (Table 2). The majority (3172) of our observations were discarded. Only the 965 "good" observations, about 23% of the total, are retained. Note that this does not mean that of the lake delineations produced, we only retained about 23%; many of the "lake observations" we reviewed had no delineations at all due to ice or cloud cover or classifier error in the form of false negatives. Of the non-empty observations (2676), we retained about 36%. Over 2676 attempted lake observations, we identified 111 cases of false-positive delineation error in lakes which were no immediate discarded as false-positive identifications and 66 cases of complete fabrication of lakes presence, giving us a total false-positive rate of 6.6%.

Where there were noticeable errors in classification which do not represent much more than a 10-20% error in total lake area, we labeled these observations as "90%" good. These classifications retain meaningful data regarding the lake margin as most of the perimeter of the lake is well-delineated. Nonetheless, we ultimately discard these "90%" observations without including them in the final dataset.

The majority of the error we observe in our lake boundary delineations is obstruction error (Table 2). Obstruction error is the error caused when some object obscures the lake surface, rendering the water land cover type undetectable. In the cryosphere environment, the most common sources of obstruction error are cloud cover and ice cover. The remaining error is error which cannot be ascribed to any partial obstruction. Some of this error (imagery error) is due to bad Landsat data or flaws of the compositing algorithm. The rest of this data is due to mistakes of the classifier.

False positives are rare, but when they do occur, it is due to one of two causes. Either the classifier becomes confused due to wet glacial ice, which spectrally resembles water, or it becomes confused due to very dark shadow on ice or proglacial debris which also resembles water's spectral characteristics. False negatives are much more common. While they resemble obstruction error, and may be partially caused by ice or cloud presence, the classifier incorrectly fails to identify water for some reason which is insufficiently explained by ice or cloud presence. This type of error is frequent and it can be responsible for lakes completely failing to be classified despite no obstruction.

## Description of Cleaned Data

From our cleaned classifications, we have a body of highly accurate data to consider. We have identified 120 ice-marginal lakes with a total of 585 observations between all of them. This gives us an average of 4.875 observations per lake over our seven-year study period. We can only theoretically observe a lake up to four times per year with our monthly composites from the four summer months (June, July, August, September), so this is a fairly poor value.

However, not all lakes are poorly-observed (Figure 10). Some lakes reach an average of up to three good observations per year. Going forward into our Monte Carlo analysis of area change

uncertainty in poorly-observed lakes, we will rely on these better-observed lakes to form the basis for our simulations.

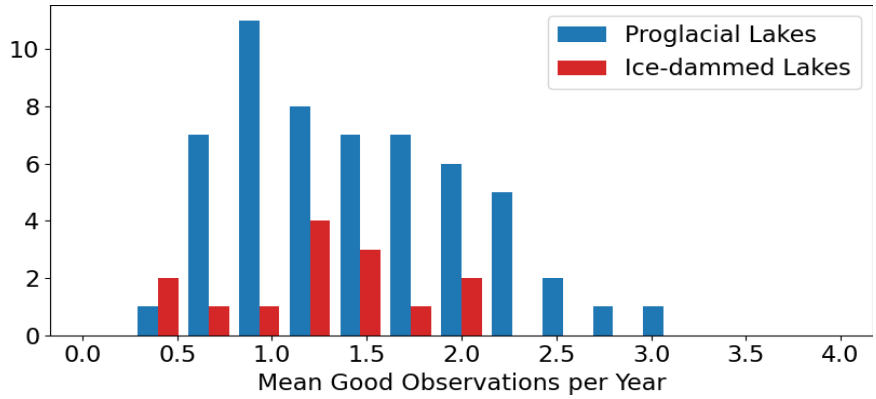



Figure 10: Histogram of observational frequency for ice-marginal lakes. This is after the manual review filtered many imperfect lake boundary delineations out.

Nevertheless, we have a database of repeated observations upon which we may perform linear regression on. We regress our 141 ice-marginal lake timeseries to obtain mean area change rates for each ice-marginal lake (Figure 11). We observe a relationship between lake area change rate and lake area. This association holds even for shrinking lakes—larger growing lakes are growing faster, but larger shrinking lakes are also shrinking faster. We find the following linear relationships that roughly describe our data:

$$\textit{Annual Area Growth}\ [\mathrm{m}^2] = 30{,}000 * \textit{Area}\ [\mathrm{km}^2]\ \text{(Gaining lakes)}$$

$$\textit{Annual Area Loss}\ [\mathrm{m}^2] = -10{,}000 * \textit{Area}\ [\mathrm{km}^2]\ \text{(Losing lakes)}$$

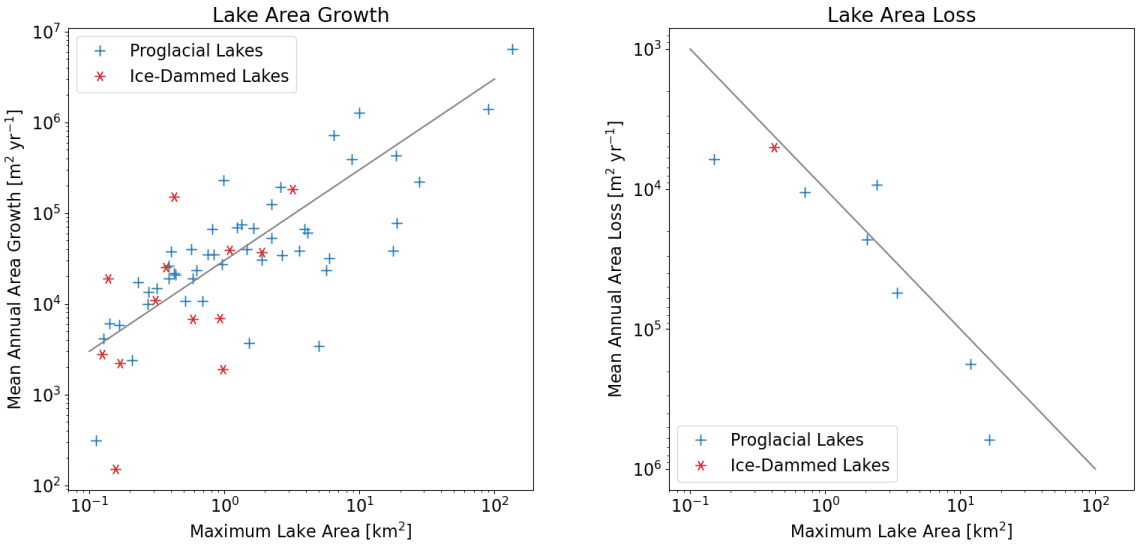



Figure 11: Lake area change rates, plotted against lake area. Note that these are log-log axes, so we must show growth and loss on separate axes.

These relationships (gray lines, Figure 11) are equivalent to saying "ice-marginal lakes which grow, grow by 3% per year" and "ice-marginal lakes which shrink, shrink by 1% per year." That

this relationship is so apparent justifies our later discussion of lake area change behavior in terms of percent per year. However, these simple relationships do not belie the variation present in the data, particularly in ice-dammed lakes. While there seems to be a separate, steeper relationship governing growing ice-dammed lakes, our data volume for these lakes is so low we do not attempt to describe them independently of proglacial lakes.

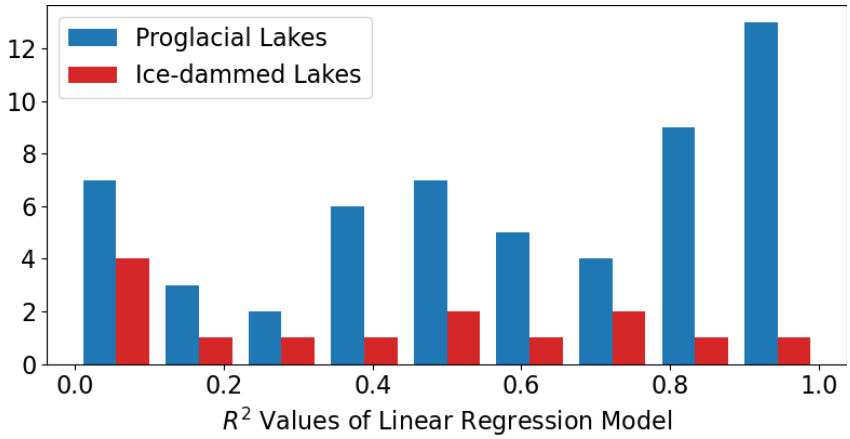



Figure 12: Histogram of quality-of-linear-fit ($R^2$) values for ice-marginal lakes.

While we obtain annual change rates from linear regression models, we may question how accurately linear regression can model ice-marginal lakes. We can use the summed square residual values (R2) to characterize the quality of these linear fits. We observe (Figure 12) there is a trimodal distribution describing quality of linear fit for proglacial lakes, with one peak nearing an $R^2$ value of 1 (indicating a perfect fit), an intermediate peak, and a third peak approaching an $R^2$ value of 0 (indicating no linear correlation). Ice-dammed lakes have a seemingly bimodal distribution, with a similar intermediate peak and a strong peak approaching an $R^2$ value of 0, but there is peak towards an $R^2$ value of 1.

## Results of Monte Carlo Simulations

We run Monte Carlo simulations for a variety of different study types. While we expect more limited studies, such as those with only two observations from consecutive years (Figure 14a), to have substantial potential error in their calculated lake area change rates, there are tactics which can decrease this potential for error. Observing lakes more than twice (Figure 14b) and observing lakes with a greater gap between observations (Figure 14c) still have inherent uncertainty, but the possibility for extreme error is reduced.

We run Monte Carlo simulations on each lake for each study type. Note that due to the limitations of our data—only a 7 year span—we cannot model studies which have more than a 5-year gap between observations, and, while we could theoretically model a study which has 7 data points each from different years, in practice, we have no more than 6 years represented significantly enough to draw meaningful conclusions. Additionally, there are other bounds—we cannot model a study of five observations with at least two years between every observation. As we reach each limit, our data volume decreases as we have fewer lakes with sufficient data to run our Monte Carlo models.

As expected, the uncertainty range decreases as the number of observations increases or as the minimum gap between observations increases (Figure 14, note x-axes). Note the histograms presented in Figure 14 are not histograms of area change rates simulated in a given Monte Carlo simulation, as illustrated in Figure 9. These are histograms of the confidence intervals derived from those histograms—each data point represents a lake which has had 100,000 Monte Carlo

simulations run. The x-axis is the width of the 95% confidence interval, reported in % area change per year.

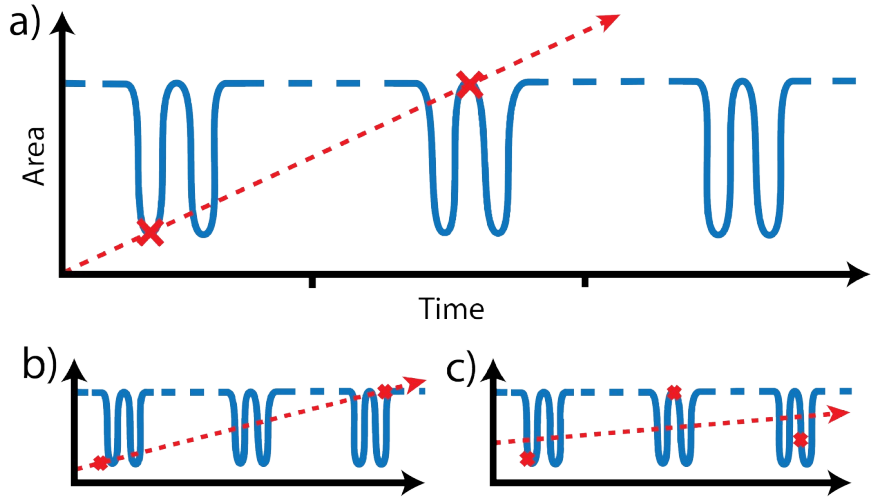



Figure 14: Conceptual models of a variable lake observed twice in consecutive years (a), with at least a one-year gap between observations (b), and three times in consecutive years (c).

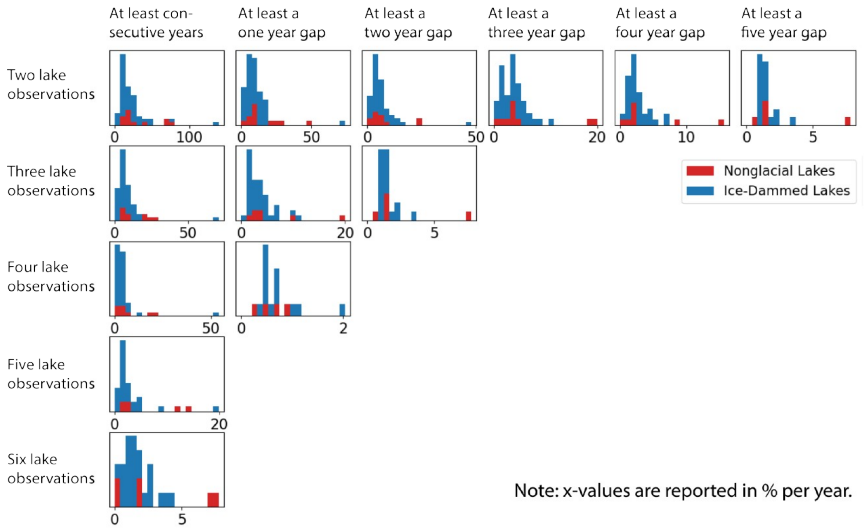



Figure 14: Histograms of 95% confidence interval widths, derived from Monte Carlo simulations of lake observations with properties indicated by x- and y-axes.

# Discussion

## Analysis of Results of the Manual Review

While the manual review of our data was performed as a methodological step to refine our data, the process of reviewing 4137 lake observations unsurprisingly proved insightful. We discuss the patterns of physical obstruction we observe in the results of the manual review.

Cloud cover is a ubiquitous problem in surface remote sensing. Its impact upon the quality of remote sensing data can be mitigated by compositing imagery together to produce cloud-free mosaics. However, as the time span of a composite is reduced, the imagery volume is also reduced, possible reducing the number of cloud-free scenes to composite. Resultingly, short-

time-scale studies such as ours will must contend with the issue of cloud cover on the data. Additionally, cloud cover can partially obscure our lakes of interest, causing detected lake area to be lower than in reality without making it immediately apparent that the variation in lake area is not due to the short-term effects we seek to study. Without proper filtering, this incorrectly low area value could be taken as a true delineation, invalidly suggesting the lake experiences rapid swings in area.

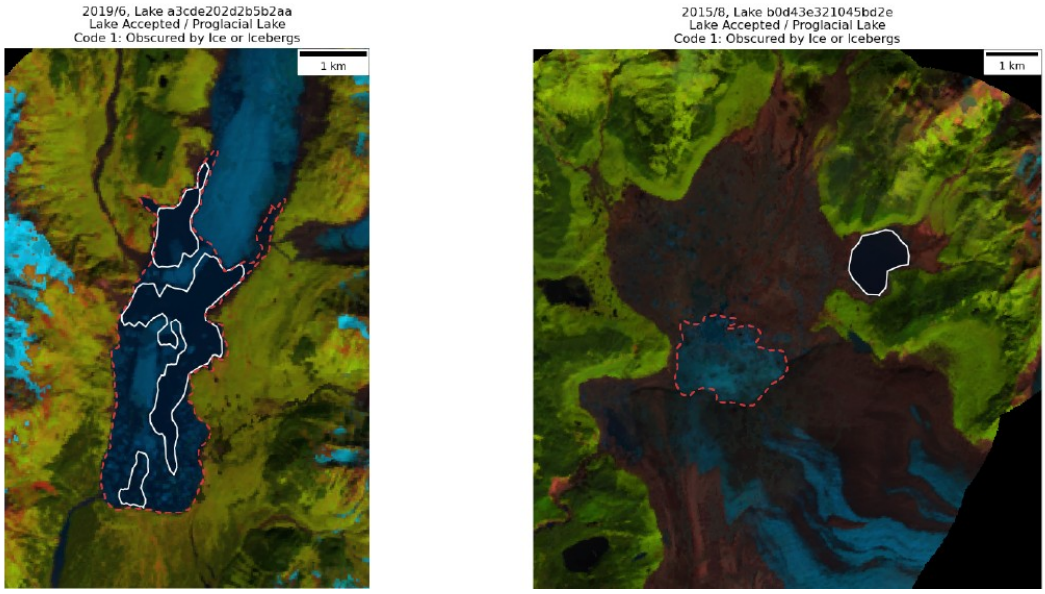



Figure 15: (Left) An iceberg-choked lake post major terminus breakup event. Red dashed line indicates where a correct delineation would be drawn. (Right) An iceberg-choked lake which experienced rapid draining. Red dashed line indicates the iceberg-choked portion of the drained lake which was unable to be automatically delineated. Green dashed line indicates the maximum fill area.

Unique to the cryosphere environment is the problem of ice cover. Besides annual lake ice and snow obscuring lakes in the winter months, we observe that ice-marginal lakes may contain icebergs which are present yearlong. Iceberg presence ranges from isolated small icebergs floating in the middle of a lake to volumes of icebergs completely choking a lake. While such a technique will confound absolute measures of lake area where islands are present, "filling the holes" in lake delineations proved an effective method for reducing the negative impact of solitary icebergs on the quality of the data.

We observe proglacial lakes growing through major glacier terminus failure (*Walder et al.*, 2006; *Tsutaki et al*., 2011; e.g.: *Chernos et al.*, 2016). This is step-wise growing behavior; major growth is associated with brief periods in which the glacier terminus faults and ruptures, exposing new lake area, followed by a multi-year periods of slow, continuous retreat. Notably, these failure events produce a much larger area of icebergs than there had been ice area prior to failure, a behavior associated in the literature with sea ice failure, not glacial mechanics in freshwater lakes (*Scambos et al.*, 2009)

Larger numbers of icebergs tend to raft together and frequently are carried by the current in a lake to accumulate along the discharge margin of the lake. For geometric reasons, it is far more difficult to ignore these icebergs. Ice obstruction occurs in nearly all lakes, but it is most frustratingly associated with two types of interesting lake behavior: major calving events (Figure 15a) and rapid lake draining (Figure 15b). Ice-related obstruction error often obscure lake growth associated with these behaviors.

We illustrate an example of a rapidly and cyclically draining lake (Figure 16). This lake clearly demonstrates the frequency and magnitude of the behavior, but it is an extreme example—

many ice-marginal lakes only partially drain, may only drain occasionally (not annually), or may experience less regular and substantial area change events but may have multiple in a year and none in others (*Carrivick and Tweed*, 2013). Lake draining commonly occurs in ice-dammed lakes (Figure 16) and is associated with outburst floods (*Post and Mayo*, 1971; *Tweed and Russell*, 1999; *Huggel et al.*, 2002; *Carrivick and Tweed*, 2013; *Falatkova et al.*, 2019), but we have observed it in proglacial lakes as well (Figure 15b) where they are partially impounded by glacial ice (*Carrivick and Tweed*, 2013).

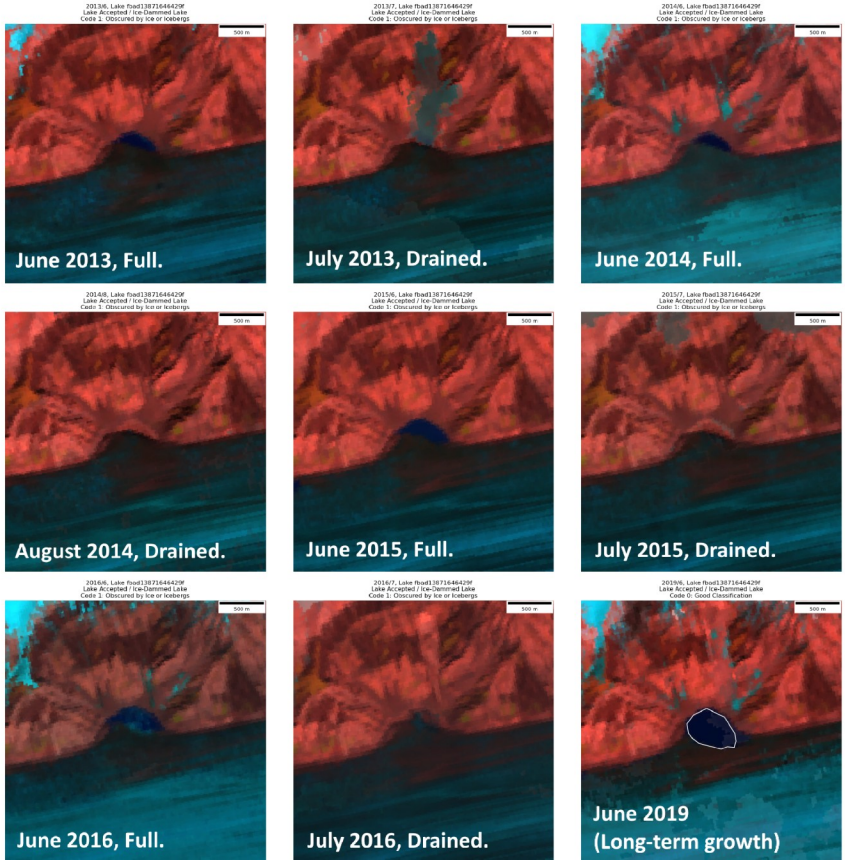



Figure 16: An example of seasonal draining and refilling of a small ice-dammed lake. The regular, cyclical behavior is apparent for several years. By 2019, the long-term growth overlaid over this behavior can be seen in the much larger full state. This lake eluded classification for many years because of its small filled size and minor iceberg presence.

When the lake drains, icebergs present in the lake collected in the drained lake area, obscuring the water surface and preventing our automated method from detecting it. Detection will not again occur until the lake is restored to a large area, distributing the icebergs widely enough that the water surface may be detected between them. However, this prevents the change in area associated with the interesting draining behavior from being correctly detected without manual review.

Overall, lake observations were roughly a quarter usable, a quarter ice-obscured, a quarter cloud-obscured, and a quarter unusable for other types of error (Figure 17). A few percent of each type of lake was given the "mostly good" code, but given that we ultimately discarded these observations, they should be considered to be a mix of the other types of error. While cloud obstruction rate was very consistent, within two percentage points of 27% across all lake types,

ice obstruction was much higher in proglacial lakes than non-ice-marginal lakes and higher yet in ice-dammed lakes.

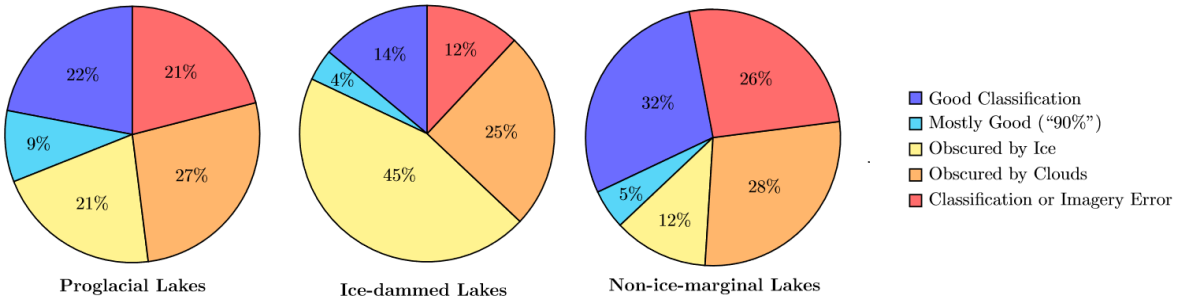



Figure 17: Summary of manual review of lake observations.

We ascribe the higher proportion of ice-obstructed observations in ice-dammed lakes to the overall colder, higher, and smaller nature of those lakes. As a result, they tend to contain more icebergs, those icebergs tend to remain unmelted longer, and those icebergs tend to obscure more of the lake at any given time. Additionally, because ice-dammed are more commonly left obscured by ice while empty or very near empty after draining.

## Discussion of Monte Carlo Simulations

While we have produced histogram summaries of the results of our Monte Carlo simulations (Figure 14), we must realize that where we have limited data (refer to Figure 10), we will struggle to characterize the uncertainty in calculated lake area change rate associated with a lower sampling rate. Our sampling rate is already very low for some lakes, and in these cases, there are very few valid ways we can resample from our data to model a given limited study—in the case of these lakes, we are a limited study.

For example, in a lake with four good datapoints, there are only six ways we could resample two observations; in a lake with three good datapoints, there are only three ways; if we have only two good observations, we are no better informed than the study we are trying to model. With so few resamplings, we may grossly underestimate the width of the 95% confidence interval for a resampling—we are victims of the same effect we wish to characterize. Thus, we must calculated the number of valid resamplings for a given resampling type and use this value to inform our analysis of the confidence intervals we calculate.

We plot 95% confidence interval widths against number of valid resamplings (Figure 18). This plotting shows us uncertainty due to the given sampling (going up on the y-axis) vs. completeness of our data for that sampling (going right on the x-axis). Consider the extreme case in which we have perfect data for every single day of the past 30 years—we would have a number of valid resamplings approaching infinity and have perfect knowledge of how the variability of lakes would affect any given hypothetical study—thus, the confidence interval we calculate for that lake would be very trustworthy.

Observe that as one reads the graph to the right (our lake data becomes more complete), there seems to exist a rising *minimum* width to the 95% confidence interval for lake resamplings. The existence of this threshold suggests that for studies of ice-marginal lakes with no further information specific to each individual lake, there still is some minimum uncertainty associated with a study of this hypothetical sort.

Further, because this threshold seems to rise and then plateau as our lake data becomes more complete, it suggests that there exists some threshold that would describe the minimum uncertainty for perfectly valid data (infinitely far to the right). If we can estimate the value of

this threshold, we can give ice-marginal studies of this sort an error bar they can use for their calculated values of lake change rates. We call this the "*minimum uncertainty threshold*". Because we know the less valid (fewer number of valid resamplings) datapoints are not giving us a meaningful representation of the uncertainty associated with low sampling, this "perfectly valid threshold" ought to form a lower bound for only those more valid data points.

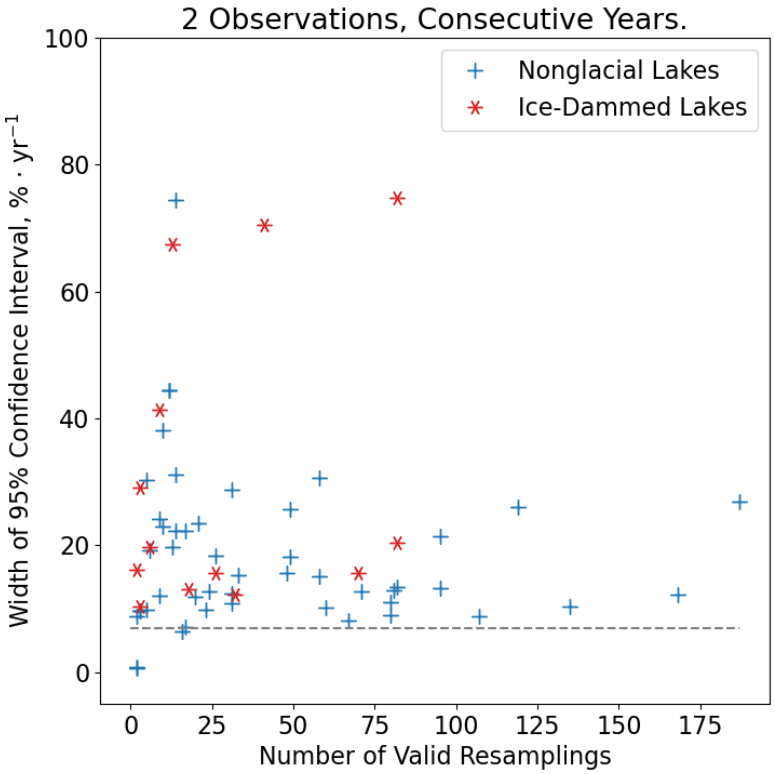



Figure 18: 95% uncertainty ranges VS. number of valid resamplings for a study of two observations in different years. These abstract axes may be thought of as "uncertainty" (y-axis) vs. "completeness (x-axis)". The 7% uncertainty threshold is marked in dashed gray.

For this particular simulated study (2 observations, data no closer than at least consecutive years), we draw this minimum uncertainty threshold at 7% lake change per year (gray line, Figure 18). This means, for a study examining lakes twice in different years, there ought to be *at least* a ±3.5% error bar on calculated percent annual lake change rates if reported with $p$=0.05. Thus, if such a study finds lakes which are changing by less than 3.5% per year, they ought to discard such findings as statistically insignificant.

We obtain the same minimum uncertainty threshold for all study types (Figure 19). We provide them as recommendations for a minimum error bar on ice-marginal lake studies of this form. There will be, and are, lakes which have more subannual variation than this minimum, and infrequently-sampling long-term change-rate studies of these lakes will be more susceptible to error caused by this short-time-scale variation.

Further, we wish to note that in studies of ice-marginal lakes, it may occur that some lakes will have better data available than other lakes. In that case, we recommend applying these

error bars to the data on a lake-by-lake basis, with a higher minimum error bar applicable if a given lake is observed more or with a larger minimum gap between observations.

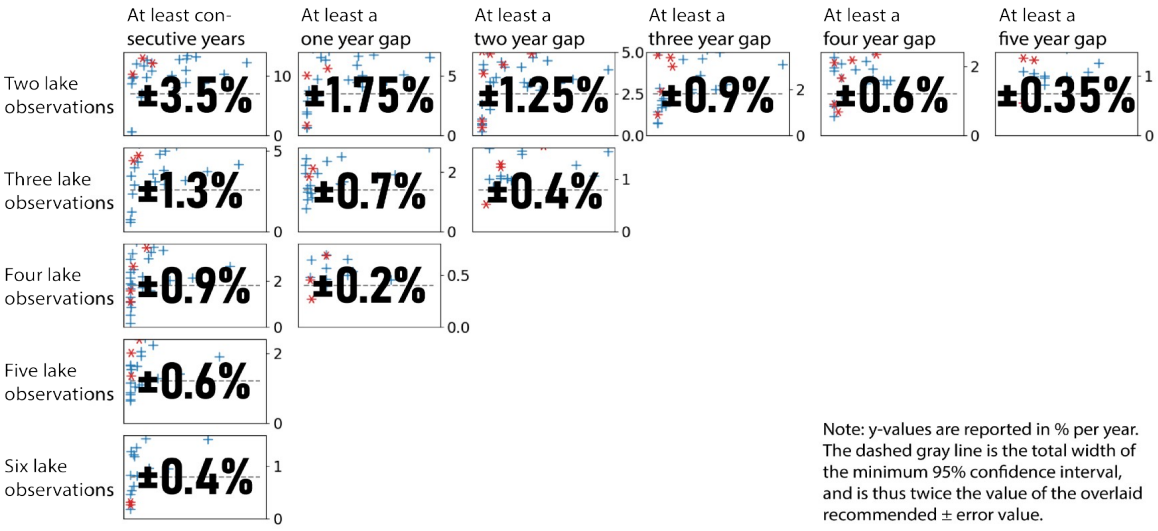



Figure 19: Minimum margin of error recommendations ($p$=0.05) for remote sensing studies of ice-marginal lakes. Number of observations per lake increases across subplots going down. Minimum gap between observations of a lake increases across subplots going right. Error bar recommendations are given in % area change per year.

# Conclusions

Ice-marginal lakes may experience large sub-annual variations in area. In ice-dammed lakes, this due is to rapid draining and refilling events, a result of the well-studied phenomenon of seasonal outburst floods. In proglacial lakes, it seems variations may be due to step-like growth from rapid major terminus breakup events. Both these short-term behaviors mean that annual observations of an ice-marginal lake are not guaranteed to be representative of the lake's long-term behaviors. From our Monte Carlo modeling, based upon our automatically-delineated, manually-filtered lake boundary data, we can make minimum recommendations for margins of error in lake area growth calculations associated with different properties of the lake observations (number and spacing).

We encourage studies of long-term behavior of ice-marginal lakes to be aware of the uncertainty introduced by the unpredictable short-term behavior of these lakes, and we offer these minimum margins of error as a first step in characterizing the difficulty in making statements regarding long-term behavior of ice-marginal lakes from limited remote-sensing data. Additionally, we want to emphasize that these minimum recommendations assume near-perfect identification of the lake. However, as discussed earlier, both clouds and ice are physical sources of obstruction error in lake identification from remote sensing. In the cryosphere environment, clouds and ice will hamper and confuse detection and accurate delineation of ice-marginal lakes, and we further encourage authors of studies of ice-marginal lakes by remote sensing to be aware of these sources of error.

We provide this study as an overview of the difficulties of remote sensing in the cryosphere environment and hope that the discussion of our methods may prove helpful for devising ways to overcome them.

# References


Bogen, J., M. Xu, and P. Kennie (2015), *The impact of pro-glacial lakes on downstream sediment delivery in Norway, Earth Surface Processes and Landforms, 40* (7), 942-952.

Carrivick, J. L., and F. S. Tweed (2013), Proglacial lakes: character, behaviour and geological importance, *Quaternary Science Reviews, 78*, 34-52.

Chernos, M., M. Koppes, and R. Moore (2016), Ablation from calving and surface melt at lake-terminating Bridge Glacier, British Columbia, 1984-2013, *The Cryosphere, 10* (1), 87-102.

Cook, S., and D. Quincey (2015), Estimating the volume of alpine glacial lakes, *Earth Surface Dynamics Discussions, 3.*

Dembczynski, K., W. Waegeman, W. Cheng, and E. Hullermeier (2011), An exact algorithm for F-measure maximization, *Advances in Neural Information Processing systems, 24*, 1404-1412.

Dorava, J. M., and A. M. Milner (2000), Role of lake regulation on glacier-fed rivers in enhancing salmon productivity: the cook inlet watershed, southcentral Alaska, USA, *Hydrological Processes, 14* (16-17), 3149-3159.

Falatkova, K., M. Sobr, A. Neureiter, W. Schoner, B. Jansky, H. Hausler, Z. Engel, and V. Benes (2019), Development of proglacial lakes and evaluation of related outburst susceptibility at the Adygine ice-debris complex, northern Tien Shan, *Earth Surface Dynamics, 7* (1).

Gong, P., and P. Howarth (1990), An assessment of some factors influencing multispectral land-cover classification, *Photogrammetric Engineering and Remote Sensing, 56* (5), 597-603.

Huggel, C., A. Kaab, W. Haeberli, P. Teysseire, and F. Paul (2002), Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss alps, *Canadian Geotechnical Journal, 39* (2), 316-330.

Jacquet, J., S. McCoy, D. Mcgrath, D. Nimick, M. Fahey, J. O'kuinghttons, B. Friesen, and J. Leidich (2017), Hydrologic and geomorphic changes resulting from episodic glacial lake outburst floods: Rio Colonia, Patagonia, Chile, *Geophysical Research Letters, 44* (2), 854-864.

Kargel, J. S., G. J. Leonard, M. P. Bishop, A. Kaab, and B. H. Raup (2014), *Global land ice measurements from space,* Springer.

Larsen, C., E. Burgess, A. Arendt, S. O'neel, A. Johnson, and C. Kienholz (2015), Surface melt dominates Alaska glacier mass balance, *Geophysical Research Letters, 42* (14), 5902-5908.

McLachlan, G. J. (1999), Mahalanobis distance, *Resonance, 4* (6), 20-26.

Otto, J.-C. (2019), Proglacial lakes in high mountain environments, in *Geomorphology of Proglacial Systems*, pp. 231-247, Springer.

Parr, W. C., and W. R. Schucany (1980), Minimum distance and robust estimation, *Journal of the American Statistical Association, 75* (371), 616-624.

Pettorelli, N. (2013), *The normalized difference vegetation index*, Oxford University Press.

Pfeer, W. T., et al. (2014), The Randolph glacier inventory: a globally complete inventory of glaciers, *Journal of Glaciology, 60* (221), 537-552.

Post, A., and L. R. Mayo (1971), *Glacier dammed lakes and outburst floods in Alaska*, US Geological Survey Washington, DC.

Scambos, T., H. A. Fricker, C.-C. Liu, J. Bohlander, J. Fastook, A. Sargent, R. Massom, and A.-M. Wu (2009), Ice shelf disintegration by plate bending and hydro-fracture: Satellite observations and model results of the 2008 Wilkins ice shelf break-ups, *Earth and Planetary Science Letters, 280* (1-4), 51-60.

Shugar, D. H., et al. (2020), Rapid worldwide growth of glacial lakes since 1990, *Nature Climate Change, 10* (10), 939-945.

Tadono, T., H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto (2014), Precise global dem generation by ALOS PRISM, ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences, 2* (4), 71.

Tsutaki, S., D. Nishimura, T. Yoshizawa, and S. Sugiyama (2011), Changes in glacier dynamics under the influence of proglacial lake formation in Rhonegletscher, Switzerland, *Annals of Glaciology, 52* (58), 31-36.

Tweed, F. S., and J. L. Carrivick (2015), Deglaciation and proglacial lakes, Geology Today, 31 (3), 96-102.

Tweed, F. S., and A. J. Russell (1999), Controls on the formation and sudden drainage of glacier-impounded lakes: implications for Jokulhlaup characteristics, *Progress in Physical Geography, 23* (1), 79-110.

Wacker, A., and D. Landgrebe (1972), Minimum distance classification in remote sensing, *LARS Technical Reports*, p. 25.

Walder, J. S., D. C. Trabant, M. Cunico, A. G. Fountain, S. P. Anderson, R. S. Anderson, and A. Malm (2006), Local response of a glacier to annual filling and drainage of an ice-marginal lake, *Journal of Glaciology, 52* (178), 440-450.

Wolfe, D. F., J. S. Kargel, and G. J. Leonard (2014), Glacier-dammed ice marginal lakes of Alaska, in *Global Land Ice Measurements from Space*, pp. 263-295, Springer.

Xu, H. (2006), Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery, *International Journal of Remote Sensing, 27* (14), 3025-3033.